Preserving Tamil Scripts:
The Way towards their Digitization, Archival and Outreach

Dr Vinodh Rajan vinodh@virtualvinodh.com

1 Introduction

The Tamil language has a written tradition that spans a little more than two millennia. In these two thousand and odd years, its writing has evolved drastically from scratching in pot shreds by potters and carvings in caves by Jain monks to being displayed in smart devices around the world. During this period, it has been written in several distinct scripts, some of which shared the label *Tamil* and others did not. Most of these scripts are now just distant memories and frozen reflections of the past.

These myriads of scripts are an integral part of the historical milieu of the Tamil land and are to be considered cultural artefacts that needs to be actively preserved for the future. However, most of the scripts are currently locked up as scanned images and cannot be properly used for diverse purposes. If someone wants to render a text portion in any of the scripts and discuss the text in its original context, it currently cannot be done (*apart from Tamil-Brahmi & Vatteluttu*) with ease. Thus, there is a need to actively preserve them by digitizing them as fonts and bring them back to life.

The article will initially explain this need to effectively preserve the scripts as digital fonts[1] and the various benefits such digitization will bestow upon us. It will then look at how this preservation can be performed, outlining the steps involved, at the same time mentioning the associated challenges. It will also briefly address the outreach efforts that are required to help the public embrace these historical scripts, without pushing them solely into the realm of specialists. Finally, it will share the personal experience of the author in digitizing and popularizing the Tamil-Brahmi script as a font.

2 A Brief Overview of Tamil Scripts

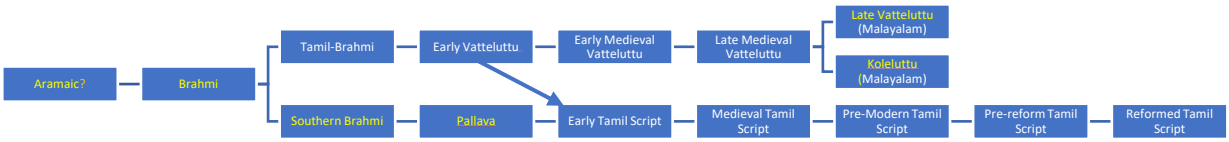


Fig 1: The evolution of Tamil scripts

The figure above shows the evolution of the various Tamil scripts from their precursors. Except for the scripts that are marked yellow, the other scripts are used to write Tamil. In fact, the language was at one-point *biscriptual* i.e. it was written in two different scripts at the same time (Siromoney, 1981). Had you hailed from the *Chera* or the *Pandya* country in the 10th century CE you would have probably used *Vatteluttu*[2], else you would have written using the *Pallava*-derived Tamil script. As opposed to popular belief, the modern Tamil script is not directly derived from Tamil-Brahmi but rather it is an adaptation of the *Pallava* script (Siromoney, 1980) with some borrowings from early *Vatteluttu.*

[1] A more appropriate term to use would be *typeface*. For the sake of avoiding jargon, *font* is used instead
[2] வட்டெழுத்து

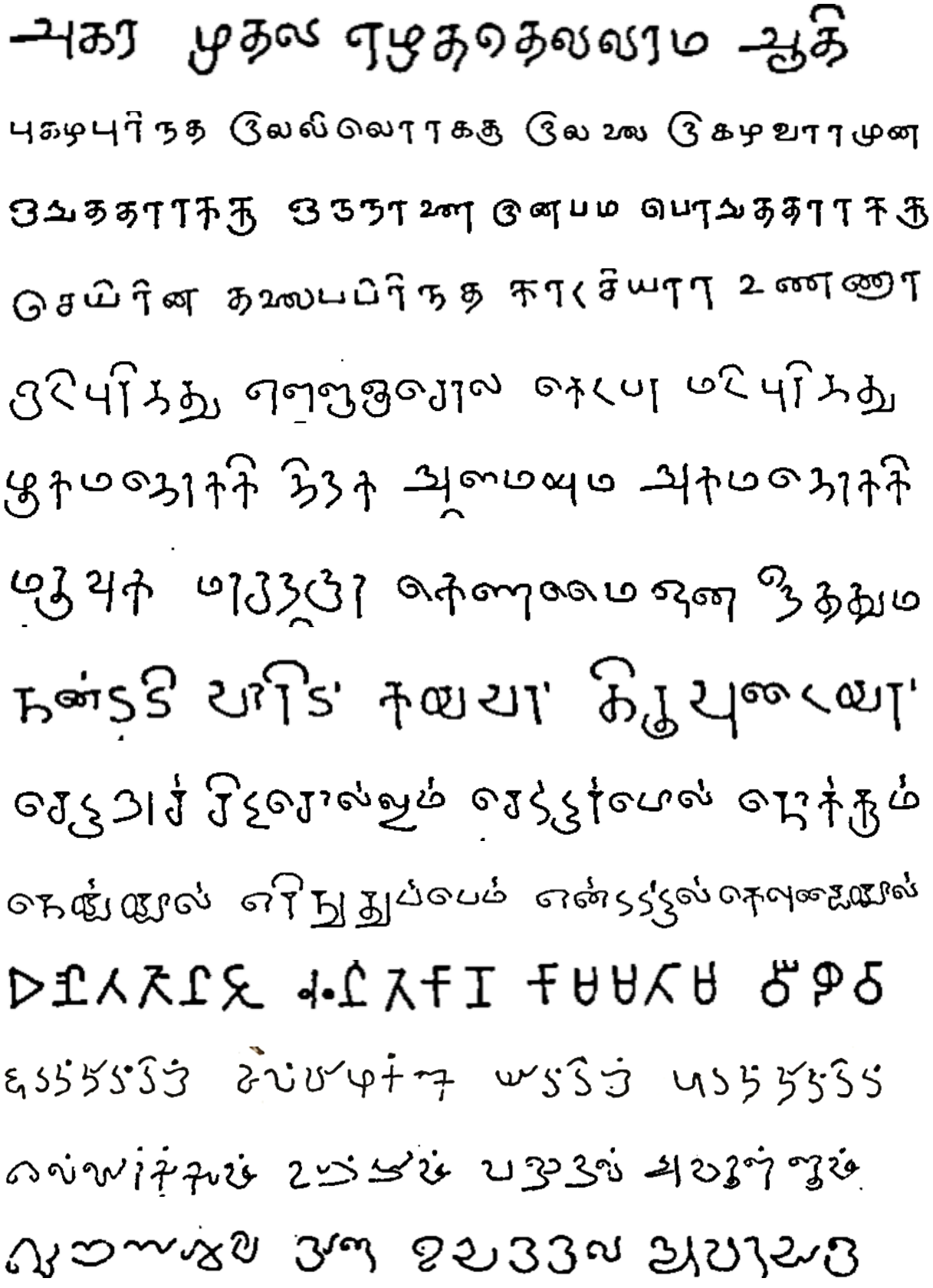


Fig 2: Samples of historical Tamil scripts from Siromoney et al. (1980) and Siromoney et al. (1981)

The scripts associated with the Tamil language can be roughly organized into three major labels: Tamil-Brahmi, *Vattelutu* and (*Pallava*) Tamil. The first two can be further categorized roughly as early, middle and late, corresponding to their evolution and distinct appearance. The (*Pallava*) Tamil script due to its longevity went through several more additional phases as shown in *fig. 1*. As a cosmopolitan language of groups that immigrated into the Tamil land and groups that emigrated out of it, Tamil has also been written with the following scripts: Telugu, Kannada, Thai Grantha and Arabic.

This wonderful diversity in Tamil scripts as shown in *fig: 2* is what needs to be preserved for the active use of the future generation.

3 Preservation as Font

It is sometimes argued that there is no need to preserve the scripts as fonts. The scripts are already preserved as scans or photographs. If we require textual representation, the content can just be rendered in the modern script. This was indeed frequently shared by many with the author during the development of the Tamil-Brahmi font. However, this does not do justice to an artefact's historical nature and are detrimental to the long-term archival of both the artefact and its script. In this context, the main arguments for preserving the scripts as fonts and its benefits is presented below.

3.1 Text Archival

Textual artefacts such as palm leaves and estampages of inscriptions are prone to destruction and decay. In the off chance such artefacts are destroyed, preserving the contents as digital textual editions in the original script ensures some form of survival of the writing in its intended shape and form.

Historical documents are also prone to interpretations. Particularly in Tamil, historical writing is rife with ambiguous sequences that need to be contextually interpreted. These are often done by experts to produce a readable text in the modern script as transcription. However, using this interpreted modern text for archival introduces a layer of subjective interpretation. The documents are no longer archived as they are. This is prevented by creating a diplomatic transcription that imitates the original material as much as possible. While diplomatic editions can be created with transliterations or the modern script, it also introduces a layer of bias. A proper diplomatic transcription should be in the original script. This will allow future researchers to access the document without any prior bias.

For instance, a reader may not realize that the transcription *cērā* 𑀘𑁂𑀭 is actually an interpretation of the original ambiguous sequence 𑀯𑁂𑀢 which in all probability could also be read as cērar 𑀘𑁂𑀭𑁆. Archiving in the original script preserves such sequences and removes interpretive bias.

All this requires the existence of a digital font for that script.

3.2 Historical Aesthetics

Nobody can deny that the content of various inscriptions and palm leaf manuscripts are of cultural and historical importance. However, the same importance ought to be given to forms of the shapes of the letters as well. A script embodies more than the underlying language that it is attempting to represent. It reflects aesthetics of that era. This nuance and impact of viewing an artefact in its original shape and form gets lost when it is rendered in the modern variant. It simply does not evoke the same emotion, mood or ambiguity.

Digitizing them and archiving them as they were written connects the audience to a bygone era and exposes them to a Tamil that existed centuries earlier.

Consider the following image of the famous Tamil-Brahmi inscription from *Jambai*.

Fig 3: Jambai Inscription (Mahadevan, 2003)

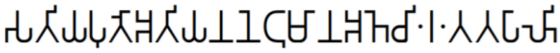
Rendering and archiving it with the apparent reading as:

ஸதியபுதோநேடுமாநஅஞுசிஈததபாளி[3]

or the intended reading as:

ஸதியபுதோநெடுமாந்அஞ்சிஈத்தப(ள்)ளி[4]

does not capture the aesthetics of the inscription as much as digitizing and archiving it as:



(Digitized text using Adinatha Tamil-Brahmi font)

This is particularly important to text passages, where an ancient word can be discussed in running text easily without resorting to inline images. One can now discuss about  of the *Sangam* period in running text with ease and comfort, all the while preserving the original aesthetics of that era.

3.3 Understanding Transmission

Rendering texts in ways they would have been originally written is essentially to understand how texts were transmitted over the ages. Some letters may have been similar in the original script of the scribe when it was written, and it could have caused confusion among copyists and, hence, resulted in transmission errors.

For instance, in Tamil-Brahmi உ */u/* and ரு */ru/* are similar and in *Vatteluttu* ட /ṭa/, ப */pa/* & வ */va/* are similar, even though the characters look visually distinct in the modern script.

Fig 4: உ /u/ and ரு /ru/ in Tamil-Brahmi

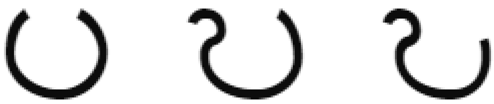


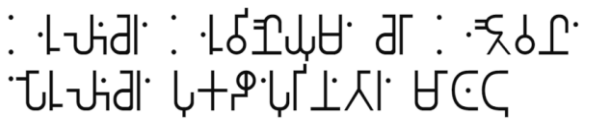
Fig 5: ட /ṭa/, ப /pa/ & வ /va/in Vatteluttu

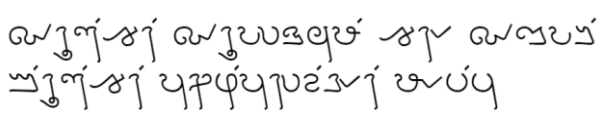
By rendering a text in various successive scripts and simulating a transmission lineage, it is possible to speculate the changes that may have occurred in texts during the transmission while copying. This will allow researchers to guess the causes for variant readings and corruptions in texts.

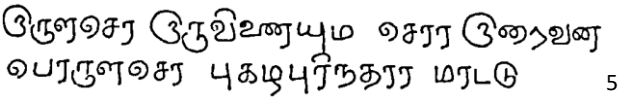
[3] satiyaputōnēṭumānañciītatapāḷi

[4] satiyaputōneṭumānañciīttapa(ḷ)ḷi

The hypothetical transmission of a *Tirukkural* couplet in multiple scripts is shown below. In each of these transmissions, there are letters that could have been confused with similar looking letters.






5

இருளசொ இருவிஜனயும சொா இறைவன
பொருளசொ புகழபுரிநதாா மாடடு

இருள்சேர் இருவினையும் சேரா இறைவன்
பொருள்சேர் புகழ்புரிந்தார் மாட்டு[6]

Existence of fonts for different Tamil scripts can enable philologists to render texts instantaneously in different scripts and perform experiments in simulated transmissions to study ancient works that have been copied across generations.

3.4 Interpreting Texts

Consider the following oft-cited pun verse:

> பத்துரதன் புத்திரனின் மித்திரனின் சத்துருவின்
> பத்தினியின் காலைவாங்கி தேய்[7]
>
> *Rub [against it] after removing the leg of the wife (tārai தாரை) of the enemy (Vālī) of the friend (Sugrīva) of the son (Rāma) of Ten-rathas (Daśaratha).*

If the *kāl ◌ா is to be removed from தாரை*, it becomes *tarai தரை* i.e. floor. Essentially, the verse is suggesting rubbing against the floor. This verse would not make any sense apart from the late medieval Tamil script, where *kāl* denotes the mark for *-ā* when occurring with consonants.

Similarly, there are verses in Tamil that make sense only if we can render them in the script that they were intended. If we are to render them in the modern script, they cannot be properly understood, as

[5] From Siromoney (1980)
[6] iruḷcēr iruviṉaiyum cērā iṟaivaṉ
poruḷcēr pukaḻpurintār māṭṭu
[7] patturataṉ puttiraṇiṉ mittiraṉiṉ catturuviṉ
pattiṉiyiṉ kālaivāṅki tēy

some of them are tied to the script of their composition. For instance, there is a verse style called *Bindumati*[8]. The 11th century Tamil grammatical compendium *Yāpparuṅkala Virutti*[9] describes it as follows (Pillai, 1998):

> பிந்துமதி என்பது எல்லா எழுத்தும் புள்ளியுடை யனவே வருவது[10]
>
> *When all letters are accompanied by a dot it is called Bindumati*

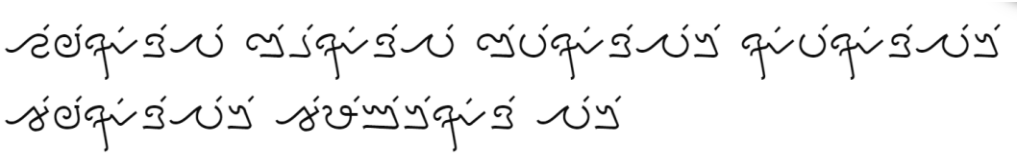
The example verse given by the compendium does not make sense if it is rendered in the modern Tamil script.

> நெய்கொண்டெ னெற்கொண்டெ னெட்கொண்டென் கொட்கொண்டென்
> செய்கொண்டென் செம்பொன்கொண் டென்[11]

However, if we are to render it using medieval conventions, where short vowels /e/ and /o/ get a dot above, the explanation of the compendium becomes visually obvious.

> நெய்கொண்டெ் னெற்கொண்டெ் னெட்கொண்டென் கொட்கொண்டென்
> செய்கொண்டென் செம்பொன்கொண் டென்

In fact, the verse makes more sense in *Vatteluttu*, where each syllable forms a contiguous glyph, unlike the *Pallava*-derived Tamil scripts where many vowel marks such as ெ and ா do not fuse.



Another instance would be the reference *ஙகர வெல்கொடியான்*[12] appearing in the *Tevāram* verse 97:16. This is glossed as *இடபக்கொடியான்*[13] *(bull-flagged-one)* and further expanded as *படுத்திருக்கும் வடிவில் இடபம் ங போன்றிருத்தலின்*[14] *(because the figure of a bull lying resembles [the letter] ங /*ṅa*/) (Sacchidanandan, -)*. Again, this probably does not make sense in modern Tamil. It may make sense if we render it in *Vatteluttu*, where the letter resembles a lying bull.

*Fig 6: ங /*ṅa*/ in Vatteluttu*

A font will allow us to render texts in the script of their composition at will and interpret them better.

---

[8] பிந்துமதி
[9] யாப்பருங்கல விருத்தி
[10] pintumati eṉpatu ellā eḻuttum puḷḷiyuṭai yaṉavē varuvatu
[11] neykoṇṭe ṉeṟkoṇṭe ṉeṭkoṇṭeṉ koṭkoṇṭeṉ
ceykoṇṭeṉ cempoṉkoṇ ṭeṉ
[12] ṅakara velkoṭiyāṉ
[13] iṭapakkoṭiyāṉ
[14] paṭuttirukkum vaṭivil iṭapam ṅa pōṉṟiruttaliṉ

4. Digitizing historical scripts as fonts

The key motivation of this section is to raise questions and uncover the tasks that need to be performed, the answers and results of which will guide the way to the various Tamil scripts' effective digitization, archival and outreach. The section does not provide definitive answers, as they require further extensive research. Any digitization attempt of a Tamil script in the future must be able to answer the questions in a convincing manner and justify their decisions.

4.1 Font Design

To summarize and echo section 3: A font allows users to create texts in a historic script of their choice with minimal effort and instantaneously. It allows creation of textual diplomatic editions and render the original inscriptions as texts but at the same preserving the original aesthetics.

However, this requires meticulous font design that truly captures the essence of the historic script. This, as one might expect, requires lots of research and painstaking referencing to published materials, original inscriptions and alphabet charts. It also involves standardizing all the variant forms and choosing the appropriate shapes for a particular script. Such a task involves a great deal of effort and time in terms of character design.

Naturally, the following questions arise:

- How do we standardize the forms?
  - Which forms to choose?
  - Which forms to reject?
- How do we capture variants?
  - Different fonts?
  - Font features?
- Do we create a font for every single era or a generic font that covers multiple eras?

One can safely assume that there is a need for at least half a dozen fonts that would reasonably cover all the eras and categories of historical Tamil scripts. History enthusiasts and volunteers may create a font or two but a consistent attempt to digitize all of the scripts requires a properly funded approach to compensate for the time and effort. A more pragmatic question would be:

- How can be this effort be funded?

4.2 Deciding the Encoding

Even if we design the shapes of the characters for a script, the next crucial step would be how to decide on the appropriate Unicode encoding for them. The existing Tamil Unicode is very efficient in expressing the modern Tamil script and is tied to the identity and behavior of that script. But this may not always hold for the historic Tamil scripts. Not only do the script grammar and behavior change but also the shapes begin to resemble modern Tamil less and less as we move back in time. At some point, this association breaks down and it may require an encoding of its own.

- How do we encode *Pallava/Chola/Nayaka* Tamil scripts?
    - Do they require separate encoding?
    - Can we unify them with the current Tamil Unicode?
    - How far can we go back without breaking the current encoding model?
    - How should the existing Tamil Unicode must be modified to accommodate historic orthographies?
- How do we encode *Vatteluttu*?

The Tamil scripts beginning from the Chola period up to the pre-modern period are quite ambiguous. This involves *pulli* being almost absent during writing and the glyphs /ra/ & vowel marker -ā (*kāl)* having the same shape. A glyph sequence such as சாா could be read as any of the following: cerā செரா, cērā சேரா, cor செர், cōr சேர், cerar செரர், cērar சேரர் based on the context. Many such sequences are common and are open to interpretation. This needs to be appropriately encoded. It is quite important for texts from manuscripts and inscriptions for their retrieval/sorting.

- How do we transfer this inherent ambiguity into the digital realm?
    - Do we need to maintain this ambiguity in terms of encoding?
    - Can we handle this at a font level and resort to higher level protocol for retreival/sorting?

Another aspect these scripts differ from the modern script is the presence of ligatures and abbreviations.

- Do they have their own character identity and, thus, require separate encoding?
- Can we handle this at the font level using font features such as ligatures?
    - If we handle it at the font level, how do we ensure their proper search and retrieval?

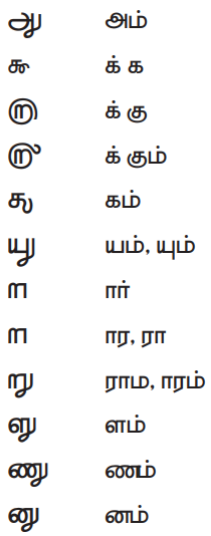


Fig 7: Some ligatures ocurring in Tamil manuscripts (Srinivas et al., 2001)

The overaching question to guide all the above is:

- How do these decisions affect script identity, character identity, text retrieval and long-term archival?

4.3 Outreach

As the famous adage goes *கடைவிரித்தும் கொள்வாரில்லை*[15] *(Even after offering, there are no takers)*, there is no point developing such intricate fonts if it is too difficult to use them and as a result there are not any takers. Creating a font is not enough. One needs to create an ecosystem that makes using the font easier that includes creating appropriate tools and software to make them accessible.

This involves creating associated tools such as Input Method Editors (IME) that allow people to compose text in the target script, converters to transform existing Tamil text into their historic predecessor scripts and, even learning resources for people who are interested in learning the scripts. Else, these will be restricted to just specialists and tech-savvy enthusiasts.

5 Tamil-Brahmi as a Case-Study

A decade ago, Tamil-Brahmi as the earliest attested Tamil script, did not have any usable digital resources available. Therefore, a collaborative effort was undertaken by the author, Udhaya Shankar, Shriramana Sharma to digitize the script and preserve the forms of earliest Tamil inscriptions in the digital realm. This was also to be used to create plain text representation of the various Tamil-Brahmi inscriptions in their original script without resorting to images or transliterations.

5.1 Adinatha Tamil-Brahmi Font

The font creation was a massive undertaking that took proper research, standardization and selection of forms. Several scholars were consulted during the design phase. It was a volunteer effort that did not receive any support and, hence, took considerable time for conceptualization to release.

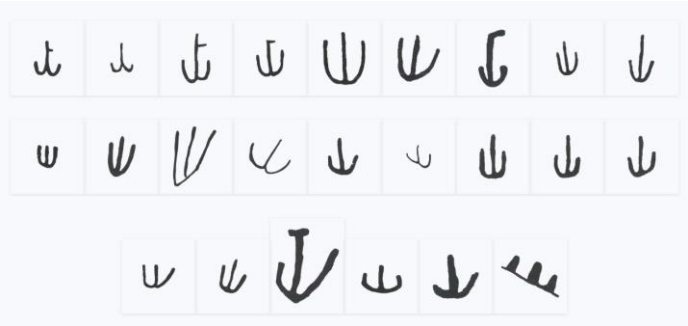


Fig 8: Variants of 𑀬 /ya/ in Tamil-Brahmi

To our advantage, Brahmi (along with the Tamil-specific characters) were already encoded in Unicode. Therefore, our font mapped the characters to the Brahmi Unicode in a one-to-one fashion. However, recently several insufficiencies in the original Brahmi Unicode to display Tamil-Brahmi characters in many modern platforms were noted. Therefore, the Unicode standard had to be updated (Rajan & Sharma, 2019) to allow properly display of the Tamil-Brahmi characters in all platforms. This resulted in the addition of 6 new Tamil-specific Brahmi characters to Unicode as shown in *fig. 11*.

The font can be downloaded from http://www.virtualvinodh.com/download/Adinatha-Tamil-Brahmi.zip

[15] kaṭaivirittum koḷvārillai

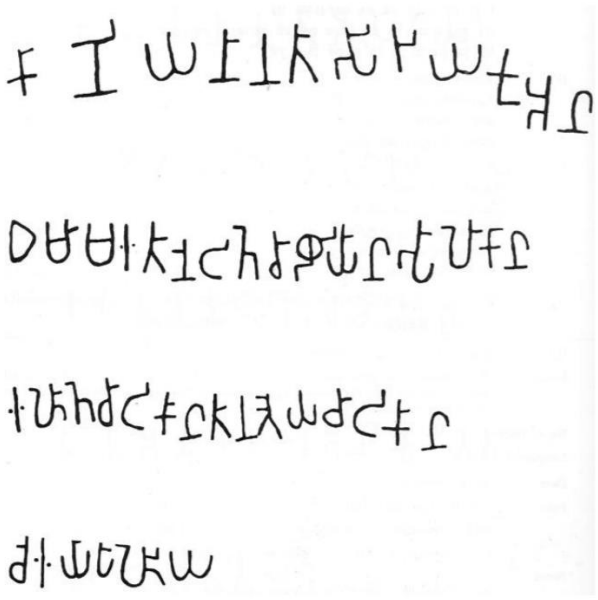


Fig 9: Mangulam Inscription (Mahadevan, 2003)

Fig 10: Digitization from Adinatha Tamil-Brahmi Font Manual (Rajan, 2012)

11070	𑁿	BRAHMI SIGN OLD TAMIL VIRAMA
11071	𑁱	BRAHMI LETTER OLD TAMIL SHORT E
11072	𑁲	BRAHMI LETTER OLD TAMIL SHORT O
11073	𑁳	BRAHMI VOWEL SIGN OLD TAMIL SHORT E
11074	𑁴	BRAHMI VOWEL SIGN OLD TAMIL SHORT O
11075	𑁵	BRAHMI LETTER OLD TAMIL LLA

Fig 11: New Tamil-specific additions to Tamil-Brahmi (Rajan & Sharma, 2019)

The discussion of the complete design process behind the Adinatha Tamil-Brahmi font can be seen in *Rajan & Sankar (2020)*.

5.2 Jinavani

Though the Adinatha Tamil-Brahmi font was released nearly a decade ago, it had very few adopters in the years that followed. Most of the early adopters were enthusiasts. The whole notion of installing a font and typing with it to compose in Tamil-Brahmi was a hurdle that prevented wide-spread adoption. Therefore, there was need to remove this. *Jinavani* (http://tamiljinavani.appspot.com) was conceived an attempt to make Tamil-Brahmi and *Vatteluttu*[16] accessible by providing a simple web/android application that requires neither font installation nor composition.

It provides a simplified interface that will allow people to convert exiting Tamil text into these scripts. To circumvent the local installation of fonts it also allows to download the rendering as images and share them in social media. Apart from this, it was made to be educational as well by providing flip cards and gamified methods to learn the scripts. This was a huge success as it allowed people to instantly get the forms. This allowed people to render their names, text passages and even complete books in both Tamil-Brahmi and *Vatteluttu*.

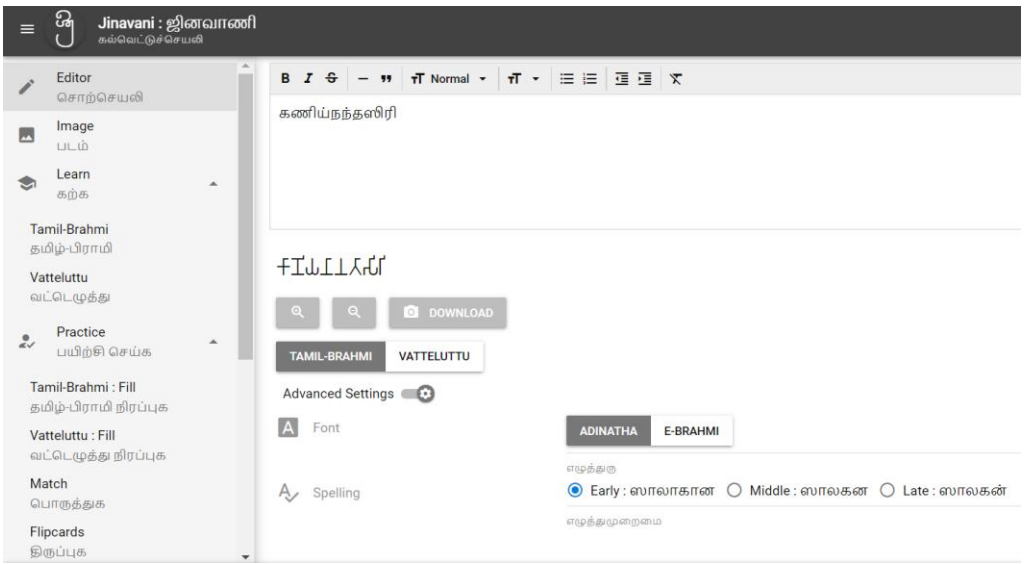


Fig 12: Jinavani Home Page

[16] The display of Vatteluttu was made by possibly through e-Vatteluttu OT as designed by Dr Elmar Kniprath

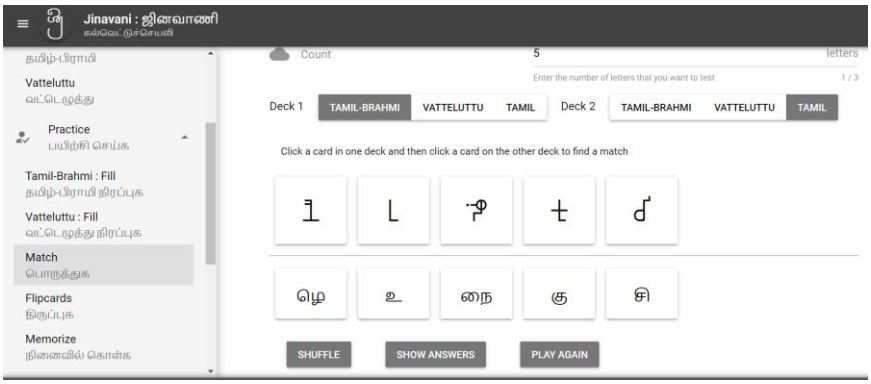


Fig 13: Jinavani Gamified Learning 1

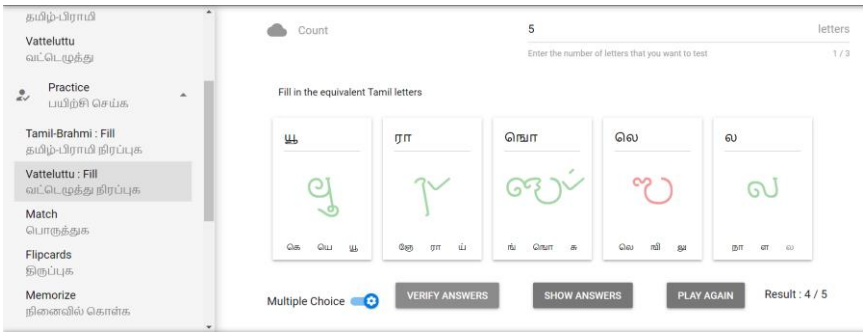


Fig 14: Jinavani Gamified Learning 2

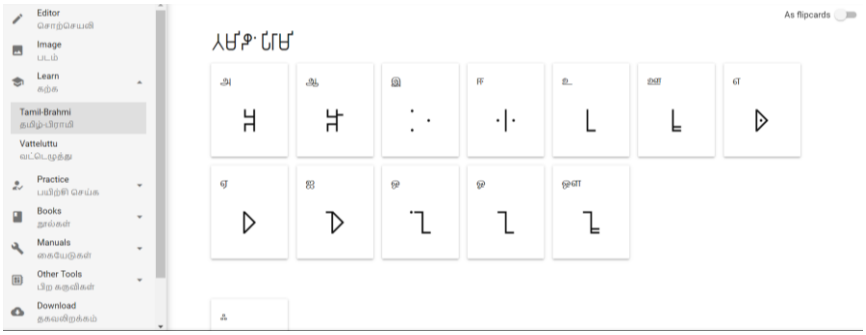


Fig 15: Jinavani Learning Cards

6 Conclusion

The road to preserving all the historic scripts used to write Tamil is a long and complicated one. This requires navigating a maze of decisions and extensive research. However, one needs to rightly recognize the need and immense value for such an effort and not brush it aside as a futile exercise. These are cultural and heritage artefacts that belong to the Tamil land and care must be taken that these are properly preserved and taken to the next generation. As much as the architectural and literary contributions of *Cheras*, *Cholas*, *Pallavas*, *Pandyas* and *Nayakas* are adored, how they wrote is also part of the Tamil heritage. Some progress has already been made in the case of Tamil-Brahmi and *Vatteluttu*. However, there is much more work to be done and with proper support and research a great deal can be further accomplished.

Acknowledgements

The author acknowledges fruitful discussions with Dr Jean-Luc Chevillard and Dr Shriramana Sharma that brought to his attention the existence of *Bindumati* and *Ṅakaravelkoṭi.* Special thanks to Dr Elmar Kniprath as well for creating the e-Vatteluttu OT font based on the *Velvikkudi* copper plates. Udhaya Sankar must also be mentioned here for providing his WIP tool that shows variations of characters in Tamil-Brahmi.

References

1. Mahadevan, I (2003): *Early Tamil epigraphy: From the Earliest Times to the Sixth Century A.D*. Chennai, India. Cre-A. (Harvard Oriental Series, Volume sixty-two.)
2. Piḷḷai, V. (1998). Yāpparuṅkalam - Paḻaiya Viruttiyuṭaṉ. Ulakat Tamiḻ Ārāycci Niṟuvaṉam, Ceṉṉai.
3. Rajan, V & Sankar, U. (2020). *Contemporizing a Historic Script: The Development of Adinatha Tamil Brahmi Typeface and its Community Impact*. Atypi 2020. Association Typographique Internationale. https://events.bizzabo.com/254216/agenda/session/418615
4. Rajan, V (2012). Adinatha Tamil-Brahmi Font Manual. http://virtualvinodh.com/pdfs/Adinatha_Tamil_Brahmi_Manual.pdf
5. Rajan, V. & Sharma, S. (2019). L2/19-402: *Proposal to Encode 6 Characters in the Brahmi Block*. L2 Registry, Unicode Consortium.
6. Sacchidanandan, M., -. Paṉṉiru Tirumuṟai Pāṭṭum Poruḷum. [online] Thevaaram.org. [Accessed 6 December 2020]. Available at: <http://www.thevaaram.org/thirumurai_1/songview.php?thiru=5&Song_idField=50970&padhi=097&startLimit=16&limitPerPage=1&sortBy=&sortOrder=DESC> (See also: https://www.ifpindia.org/digitaldb/site/digital_tevaram/INDEX.HTM)
7. Siromoney, G., Govindaraaju, S. & Chandrasekaran, M. (1980). *Tirukkural in Ancient Scripts*. Tambaram. Department of Statistics, Madras Christian College.
8. Siromoney, G., Govindaraaju, S., Chandrasekaran, M. & Chandrasekaran, S. (1981). *Vaṭṭeḻuttil Tirukkuṟaḷ*. Tambaram, Department of Statistics, Madras Christian College.
9. Srinivas, M.D., Paramasivam, T.G. & Pushkala, T. (2001). *Thirupporur and Vadakkuppattu: Eighteenth Century Locality Accounts*. Centre for Policy Studies, Chennai.
10. Visalakshy, P. (2003). The Grantha Script. Thiruvananthapuram: Dravidian Languages Association.